



Sparrow, J. M., Grzeda, M. T., Frost, N. A., Johnston, R. L., Liu, C. S. C., Edwards, L., Loose, A., & Donovan, J. L. (2018). Cat-PROM5: a brief psychometrically robust self-report questionnaire instrument for cataract surgery. *Eye*, 32(4), 796-805.
<https://doi.org/10.1038/eye.2018.1>

Publisher's PDF, also known as Version of record

License (if available):
CC BY-NC-ND

Link to published version (if available):
[10.1038/eye.2018.1](https://doi.org/10.1038/eye.2018.1)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Cat-PROM5: a brief psychometrically robust self-report questionnaire instrument for cataract surgery

JM Sparrow^{1,2}, MT Grzeda^{1,2}, NA Frost³,
RL Johnston⁴, CS C Liu^{5,6}, L Edwards¹, A Loose¹
and JL Donovan²

Abstract

Purpose To develop a short, psychometrically robust and responsive cataract patient reported outcome measure suitable for use in high-volume surgical environments.

Methods A prospective study in which participants completed development versions of questionnaires exploring the quality of their eyesight using items harvested from two existing United Kingdom developed parent questionnaires. Participants were 822 patients awaiting cataract surgery recruited from 4 cataract surgical centres based in the UK.

Exclusion criteria were other visually significant comorbidities and age <50 years. An iterative multi-stage process of evaluation using Rasch and factor analyses with sequential item reduction was undertaken.

Results A definitive item set of just five items delivered performance in accordance with the requirements of the Rasch model: no threshold disordering, no misfitting items, Rasch-based reliability 0.90, person separation 2.98, Cronbach's α 0.89, good targeting of questions to patients with cataract with pre-operative item mean -0.41 logits and absence of significant floor or ceiling effects, minor deviations of item invariance, and confirmed unidimensionality. The test-re-test repeatability intra-class correlation coefficient was 0.89 with excellent responsiveness to surgery, Cohen's $d = 1.45$ SD. Rasch calibration values are provided for Cat-PROM5 users.

Conclusions A psychometrically robust and highly responsive five-item cataract surgery patient reported outcome measure has been developed, which is suitable for use in high-volume cataract surgical services.

Eye (2018) 32, 796–805; doi:10.1038/eye.2018.1; published online 9 March 2018

Introduction

Cataract surgery is one of the most frequently undertaken surgical procedures globally.^{1,2} Traditionally, monocular visual acuity has been used to assess pre-operative need for surgery and post-operative success. The inadequacies of this approach have been widely recognised³ and in recent times patient reported outcome measures (PROMs) have attracted greater emphasis with a plethora of instruments being offered for use in cataract.^{3,4} Despite the existence of available questionnaires there have been recent high level calls for better PROM instruments for cataract surgery in the NHS,⁵ including a 2017 high priority research recommendation from the National Institute for Health and Care Excellence (NICE).⁶ Early instruments were developed using Classical Test Theory, with many having subsequently been re-evaluated using modern item-response-based statistical techniques, in particular, Rasch analysis.⁴ This approach to vision-related self-report questionnaires provides for development of a unidimensional instrument capable of measuring an underlying 'latent trait' of visual difficulties. Rasch analysis allows sets of questions to be analysed to reveal whether a single or multiple measurement constructs are being addressed by the questions. Previous studies have adopted an approach where items from existing questionnaires are grouped into unidimensional subscales, each of which measures a slightly different construct.^{7,8} For a valid assessment of dimensionality a certain number of items are required, typically around 10 questions being deemed sufficient, although as few as 3 or 4 questions have been analysed in this way to confirm or refute unidimensionality.^{7,9} Item-banking of questions

¹Bristol Eye Hospital, Lower Maudlin Street, Bristol, UK

²Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

³Torbay Hospital, Lawes Bridge, Torquay, UK

⁴Gloucestershire Hospitals NHS Foundation Trust, Cheltenham General Hospital, Cheltenham, UK

⁵Sussex Eye Hospital, Brighton, UK

⁶Brighton and Sussex Medical School, Brighton, UK

It is with deep regret that we note the death of our coauthor, friend, and colleague Robert L Johnston, who sadly died in September 2016.

Correspondence: J Sparrow, Bristol Eye Hospital, Lower Maudlin Street, Bristol, BS1 2LX, UK
E-mail: John.Sparrow@doctors.org.uk

Received: 1 July 2017
Accepted in revised form: 21 December 2017
Published online: 9 March 2018

can provide a useful research tool^{10,11} and may be applicable where large item sets are available. There may, however, be disadvantages to item-banking, the same questions, or same number of questions, are not completed by patients, it does not enable fixed scoring systems, is less suitable for specialised specific latent traits, and prevents returning to earlier questions to amend responses. The approach used here was to select the smallest number of items compatible with good psychometric performance, an approach which ensures that the best-performing items are used on each occasion. The small number of fixed items maintains flexibility by allowing for either pen and paper completion or electronic entry of responses by patients themselves. To be of practical value in high volume cataract surgical settings it is critically important for questionnaire instruments to be brief. Psychometrically, a trade-off exists between questionnaire length and performance, including responsiveness to surgical intervention, making questionnaire design and item selection paramount. In this paper, the development of Cat-PROM5, a very brief five-item cataract patient reported outcome measure is described, illustrating performance similar to current 'best of class' longer instruments.

Materials and methods

Study design

The setting of the study was across 4 cataract surgical centres (Bristol, Torbay, Cheltenham, Brighton) in the English National Health Service (NHS). Questions were harvested from 2 existing United Kingdom (UK) developed questionnaires, the Visual Symptoms and Quality of life questionnaire (VSQ)¹² originally developed for a randomized trial of second eye cataract surgery¹³ and the Vision Core Module 1 (VCM1)¹⁴ originally developed as a generic Vision Related Quality of Life (VR-QoL) questionnaire. The original items were separately generated through 40 (VSQ) and 38 (VCM1) in-depth interviews with patients, with subsequent operationalisation involving a further 58 patients (VCM1). Building on this earlier work, the full set of VSQ items were reviewed and those deemed too complex and/or of low applicability excluded. Ten VSQ items were retained and re-operationalised together with 10 VCM1 items and an additional general vision question, giving an initial set of 21 items for evaluation. These items included three theoretical constructs related to self-reported issues with vision: (1) visual functioning (also known as visual disability, or activity limitations); (2) visual symptoms; (3) emotional impacts of vision.

Rather than attempting to impose an *a priori* theoretical subscale classification onto questionnaire items, a data

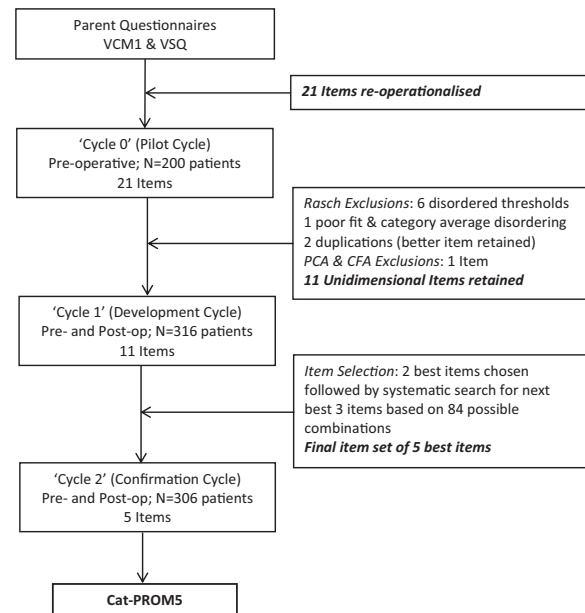


Figure 1 Flow chart for Cat-PROM5 development study. People awaiting cataract surgery without other visually significant comorbidities in either eye were recruited following informed consent.

(patient) led iterative multistage design was employed to simply eliminate subscales. This included three separate data collection cycles as outlined in Figure 1. For the initial pilot or 'Cycle 0', baseline pre-operative questionnaire completions were analysed by Rasch followed by Factor Analyses to exclude disordered and misfitting items and assess dimensionality. Item reduction continued until a unidimensional item set had been achieved. At this stage, based on the pilot data, the retained unidimensional items 'moved together' indicating that a single construct or 'aspect of vision' was being measured collectively by these items. At the next stage, 'Cycle 1', both pre-operative and post-operative questionnaire completions were analysed. Psychometric performance of retained items was checked to confirm performance, including unidimensionality, and their responsiveness to surgical intervention was estimated. Having eliminated items belonging to constructs other than the central focus of the item set, the retained items were deemed to measure a single construct, which we describe as visual difficulty related to cataract. Further item reduction using a comprehensive assessment process resulted in selection of a definitive five-item set. In the final confirmation stage, 'Cycle 2', performance of the selected definitive five-item set was re-evaluated using a further sample of pre- and post-operative questionnaire completions. As part of 'Cycle 2' a 1 in 5 random subsample of participants made a second pre-operative questionnaire completion at least 2 weeks following the first to provide for a test-re-test analysis. Finally, data for

Table 1 Psychometric properties of the scale and criteria for acceptability

<i>Psychometric property</i>	<i>Aim definition</i>	<i>Criteria for acceptability for unidimensional scale</i>
Valid measurement model	To identify a pool of items which effectively measure the concept of visual difficulty. To remove items that do not fit the assumed criteria of a unidimensional measure	<p>Applied to Rasch modelling: Rasch-Andrich thresholds ordered as expected; Mean Square fit statistics: Outfit/Infit within 0.7–1.3^a (ref. 24) Point-measure correlation $\geq 0.4^b$; Category averages ordered as expected; Unidimensionality: PCA highest eigenvalue of residual correlation matrix < 2.0 (ref. 25); Item invariance: DIF < 0.43 logits regarded as small/negligible, $0.43 \leq DIF < 0.64$ slight or moderate, $DIF \geq 0.64$ high and significant^(ref. 23)</p> <p>Applied to Exploratory Factor Analysis: Kaiser–Guttman criterion: one eigenvalue of the correlation matrix > 1.0 Cattel’s scree test: one eigenvalue above inflection point; factor loadings above 0.30^(ref. 26)</p> <p>Applied to Confirmatory Factor Analysis:^c RMSEA ≤ 0.09 (refs. 27–30), CFI > 0.95 (ref. 31), factor loading p’s < 0.05</p>
<i>Reliability</i>		
Precision	The reliability indexes assessing the precision of the measure. Two indexes were of our interest: (a) Rasch-based reliability is the share of the ‘true’ variance in the total observed variance of the measure. (b) Person Separation index; the ratio of the reliable (‘true’) variation in measure to the variation stemming from the random noise. Both indices serve as Rasch equivalents to traditional reliability indices (Cronbach’s alpha.)	Rasch based reliability index: ≥ 0.7 acceptable reliability ≥ 0.8 good; ≥ 0.9 excellent Person separation > 2.0
Test–re-test reliability	Confirmation that the items of the scale return stable results assessed by administering the same questions repeatedly to the same patients in the absence of a change in clinical status.	Intraclass Correlation > 0.70 Cohen’s Kappa: > 0.5 moderate > 0.6 good
Scale responsiveness	Evidence that the scale is sensitive to surgical intervention for cataract (mean change in Logits divided by the standard deviation computed for the all pre- and post-operative patients combined).	Effect size: Moderate > 0.50 SD Large > 0.80 SD
<i>External criteria</i>		
Discriminative validity	Evidence that the measure of visual difficulty is not simply a repetition of an existing clinical measure, that is, the instrument should capture information relevant to the wider experience of a person’s vision.	Low (< 0.3) correlation with visual acuity (LogMAR)
Convergent validity	Evidence that the measure is highly correlated with other similar visual difficulty PROMs.	High (≥ 0.7) correlation with Catquest-9SF

^a Outfit/Infit statistics < 0.7 suggest item redundancy, > 1.3 indicates poorly fitting items.²⁴ ^b Caution should be exercised before removing items located towards either end of the scale as these have lower correlations but may enhance precision towards the scale extremities. ^c Root-mean-square-error-of-approximation (RMSEA)^{27,28,30} with threshold 0.09 stricter than usual 0.08 in view of categorical variables, sample < 250 and type I error near 5% (ref. 29) and comparative fit index (CFI).

the definitive five-item set from all Cycles were aggregated for a combined analysis, which included calibration of the questionnaire items.

People with age related cataract who were awaiting first or second eye cataract surgery at participating centres were potentially eligible for recruitment. Inclusion criteria were: age 50 years or older, ability to understand and complete development versions of Cat-PROM and Catquest-9SF in English, willingness to participate and exclusion criteria were: visually significant ocular or systemic comorbidity, for example, advanced age-related or diabetic maculopathy, significant amblyopia (VA worse than $6/12 = 0.3\text{LogMAR}$), gross visual field loss (any cause) or any other visually significant ocular or systemic comorbidity that in the opinion of the local principal investigator rendered the patient unsuitable for the study. These criteria were used to recruit typical NHS patients approaching cataract surgery and to avoid possible confusion between vision issues due to cataract and non-cataract comorbidities. As a precaution, and as reported elsewhere,¹⁵ a separate qualitative study was undertaken with people who had both cataract and other visually significant comorbidities to check that these did not cause serious difficulties with the use of the questionnaire in individuals with both cataract and other causes of vision loss. Data were transcribed to a purpose built study database at study sites, with regular source data verification to assure data quality. The study was conducted in compliance with all applicable regulatory requirements (ethics ref:13/NW/0616).

Rasch modelling

Although Rasch proposed his model as a solution for measurement problems specific to educational testing the ideas underlying this model have been adopted as a tool for construction and validation of whole-person concepts such as attitudes, symptoms, perceptions, and (dis)abilities.^{16–18} The method provides an estimation mechanism for conversion of ordinal questionnaire data into an interval measure which conforms to the axioms of fundamental measurement, more familiar in the physical sciences.^{16,19} This measure takes the form of the Rasch continuum in units of logits, positioning both respondents and items (and their categories) onto the same underlying latent scale, in this case that of self-reported issues with vision due to cataract.

The process of Rasch scaling amounts to a series of iterative procedures testing whether fundamental assumptions of the model hold for a particular set of items or questions, with sequential exclusions. When generating the Rasch parameters, to avoid violation of the underlying assumptions of the model we used only a single completion per person, these being randomly selected as

either pre- or post-operative completions, but never both. Since the question structures and rating categories varied, analysis using the Rasch partial Credit Model (PCM)²⁰ was appropriate and this was complimented by supplementary Exploratory and Confirmatory Factor Analyses using polychoric correlations (EFA and CFA). The combination of Rasch and Factor Analyses provide a comprehensive mechanism for assessing dimensionality, that is, checking that all the questions relate to the same underlying construct, in this case visual difficulty related to cataract. Item invariance was checked through differential item functioning (DIF) by analysis of patient data split using 8 sets of criteria, with attention paid to both the statistical significance and magnitude of observed contrasts. The purpose of DIF analysis is to test whether individual questions are used in the same way or differently by individuals belonging to identifiable subgroups, for example, male *vs* female or younger *vs* older. The list of analytical parameters deployed in the development process, along with acceptance / rejection criteria are summarised in Table 1.

To assess the scale's responsiveness to surgical intervention we considered the pre- to post-operative mean differences in Logits and Cohen's *d*, the latter calculated by two methods (to facilitate comparisons with other studies), firstly using the theoretically more sound pre-operative baseline SD and secondly using the traditional pooled pre- and post-operative SD.

Results

Study participants

Across all three cycles of the study there were 822 participants with analysable data on 1266 completed questionnaires. Demographic and other information on study participants is given in Table 2.

'Cycle 0' or pilot cycle

From the initial item set of 21 questions, items were excluded iteratively, at each successive step the most problematic item being removed prior to Rasch PCM reanalysis. Following exclusions (Figure 1), 12 items remained for which the fundamental Rasch assumptions held. Principal Component Analysis PCA on residual variance gave a borderline dimensionality result, and this along with a high residual correlation between two items suggested the possibility of two sub-dimensions. CFA confirmed a need to exclude a further item, following which all analysis parameters were satisfactory. Eleven unidimensional items were taken forward to the next analysis cycle (see online Supplementary Table S1 for item

Table 2 Sociodemographic characteristics of participants

	'Cycle 0' or Pilot (baseline only) N = 200	'Cycle 1' (baseline and follow-up) N = 316	'Cycle 2' (baseline and follow-up) N = 306	All cycles N = 822
Age median (1 st Qr; 3rd Qr)	76 (70; 81)	76 (70; 82)	76 (70; 82)	76 (70; 82)
Gender, M:F (N, Col %)	71:129; 35.5%:64.5%	131:183; 41.5%:57.9%	136: 170; 44.4%:55.6%	338:482; 40.9%:58.4%
Missing	0; 0.0%	2; 0.6%	0; 0.0%	2; 0.2%
Side R:L (N, Col %)	107:73; 53.5%:36.5%	168:145; 53.2%:45.9%	162:144; 52.9%:47.1%	437:362; 53.2%:44.0%
Missing	20; 10.0%	3; 0.9%	0; 0.0%	23; 2.8%
Eye 1st:2nd (N, Col %)	154:43; 77.0%:21.5%	229:84; 72.5%:26.6%	169:137; 55.2%; 44.8%	552:264; 67.2%:32.1%
Missing	3; 1.5%	3; 0.9%	0; 0.0%	6; 0.7%
SES ^a (N, Col %)				
Q1	57; 28.5%	70; 22.2%	70; 22.9%	197; 24.0%
Q2	41; 20.5%	75; 23.7%	67; 21.9%	183; 22.3%
Q3	36; 18.0%	71; 22.5%	80; 26.1%	187; 22.7%
Q4	45; 22.5%	56; 17.7%	51; 16.7%	152; 18.5%
Q5	16; 8.0%	22; 7.0%	29; 9.5%	67; 8.2%
SES missing	5; 2.5%	22; 7.0%	9; 2.9%	36; 4.4%
Site (N, Col %)				
Bristol	196; 98.0%	107; 33.9%	93; 30.4%	396; 48.2%
Torbay	4; 2.0%	78; 24.7%	47; 15.4%	129; 15.7%
Cheltenham	—	79; 25.0%	81; 26.5%	160; 16.7%
Brighton	—	52; 16.5%	85; 27.8%	137; 19.5%

^a SES—Index of Multiple Deprivation.

descriptions and Supplementary Table S2 for Rasch parameters).

Cycle 1

Patients in Cycle 1 completed the reduced Cat-PROM questionnaire pre- and again a few weeks post-operatively. The results from Cycle 1 in general confirmed that the set of 11 items were appropriately selected with no reversed thresholds and acceptable Rasch parameters (Supplementary Table S2). DIF analysis returned only minor drifts from the specified limits. The mean self-reported visual difficulty on the Rasch scale changed between pre- and post-operatively by -2.16 logits, from -0.66 to -2.88 . The standardized effect size (Cohen's d)²¹ was -1.62 SD (pre-op SD), and -1.02 SD (pre- and post-op pooled SD). The 11 items were confirmed as a well-performing unidimensional scale measuring visual difficulty related to cataract.

Since the objective was to develop a short and responsive questionnaire suitable for high-volume cataract surgical services, the relative performance of individual items and subsets of items was considered. Preliminary probing indicated that when the item set was reduced below five items the performance, based on Rasch parameters, dropped unacceptably, identifying a five-item set as the preferred size. On a range of considerations VSQ_Overall and VCM1_Interfere stood

out as the best two candidates and it was decided that they should be included in a final item set. To search out the best subset of 5 items every possible combination of 5 items from the pool of 11 was generated, with the constraint that each subset should include VSQ_Overall and VCM1_Interfere. The 84 possible subsets were separately Rasch analysed. Through a comprehensive selection process that included assessment of Rasch performance parameters, responsiveness to surgery, patient preferences and expert opinion, the remaining three items were chosen with VCM1_Interfere, VSQ_Overall, VSQ_Reading, VSQ_Doing, VSQ_Bad_Eye being the optimum choice for the final five-item set.

Cycle 2

As the final stage of the Cat-PROM5 scale validation process, Cycle 2 was designed to check the performance of the definitive five items chosen. Rasch indices for the fresh data were similar to those from Cycle 1 and generally satisfactory (Table 3, reversed category averages of the two extreme categories of VSQ_Overall were explained by the fact that there were only 3 endorsements of the final category in this sample. There were no serious DIF problems).

On average pre- to post-operative scores changed by -3.16 logits, corresponding with a standardized effect size (Cohen's d) of -1.52 SD and -1.11 SD by the two

Table 3 Psychometric performance of the Cat-PROM5 items for the pre- and post-operative 'Cycle 2' and for all Cycles combined. (Items ordered from low to high visual difficulty from above down)

Cat-PROM5 Items	Cycle 2 Baseline and follow-up (N = 255)		Cycle 2 Test-re-test (N = 53)		Pilot and Cycles 1, 2 Combined Baseline and follow-up (N = 735†; Pilot = 200; Cy1 = 280; Cy2 = 255)					
	Measure ^a (SE)	Infit MnSq	Outfit MnSq	Point measure correlation	Linear Weighted Kappa	Quadratic Weighted Kappa	Measure ^a (SE)	Infit MnSq	Outfit MnSq	Point
measure										
Correlation										
VSQ_Bad Eye	-0.92 (0.12)	0.91	0.91	0.81	0.53	0.69	-1.00 (0.07)	1.13	1.14	0.78
VSQ_Overall	-0.68 (0.10)	1.10	1.06	0.85	0.63	0.73	-0.66 (0.06)	1.01	1.00	0.87
VSQ_Reading	-0.02 (0.11)	1.22	1.08	0.80	0.54	0.69	0.11 (0.06)	1.10	1.08	0.81
VCM1_Interfere	0.04 (0.10)	0.87	0.90	0.86	0.53	0.69	0.19 (0.06)	0.84	0.88	0.86
VSQ_Doing	1.58 (0.14)	0.88	0.84	0.74	0.57	0.66	1.36 (0.08)	0.88	0.85	0.76
Model Indices										
	Rasch-based reliability 0.88; Person separation: 2.66; Cronbach's α : 0.88; Cohen's d : -1.52 ^b or -1.11 ^c				Rasch Person Measure ICC for 5-Item Scale		Rasch-based reliability 0.90; Person separation: 2.98; Cronbach's α : 0.89; Cohen's d : -1.45 ^a or -1.09 ^c			
	Residual eigenvalues 1.5; 2nd:1.3; 3rd:1.2; 4th:0.1; 5th:<0.1				0.89		Residual eigenvalues: 1.5; 2nd:1.3; 3rd:1.2; 4th:1.0; 5th:<0.1			

†The number of patients in the validation sample is 735 which includes all from Pilot, and those sampled and available from Cycle 1 and 2. If a patient from either of these later cycles was randomly selected to contribute their post-operative measurement but this was missing, then they were dropped from the model generation group. All patients were represented in subsequent computations. The parameters of Rasch models were estimated by Joint Maximum Likelihood methods using WINSTEPS v3.72.3 software. ^a Measure represents the Item Location, also called the item difficulty (it is the average of the Rasch-Andrich model thresholds). ^b Computed with SD of baseline measures as the denominator. ^c Computed with SD of total sample as the denominator N for the Rasch performance parameters relates to the sample size used to generate the Rasch model.

methods. Test-re-test reliability on a 1 in 5 random sample of 53 pre-operative patients indicated acceptable quadratic weighted Kappa for items (0.66–0.73), and an excellent intra-class correlation coefficient for the person measures (logits) of 0.89 (Table 3).

Final calibration

In order to enhance the precision of the calibration exercise the responses to the definitive set of five items was aggregated from all study cycles. The psychometric

performance for the combined data was in line with Rasch model expectations. Figure 2 shows the distribution of item category thresholds against the distribution of patient's measures, illustrating that Cat-PROM5 is well targeted, with no serious ceiling or floor effects. DIF analysis did not indicate major problems with invariance of item difficulties across eight separate patient groupings as illustrated in Supplementary Figure S1.

The 'Overall Vision' item shift was 'slight to moderate' and in the same direction for the pre-vs. post-operative split (DIF=0.62) and the 1st *vs* 2nd eye surgery split

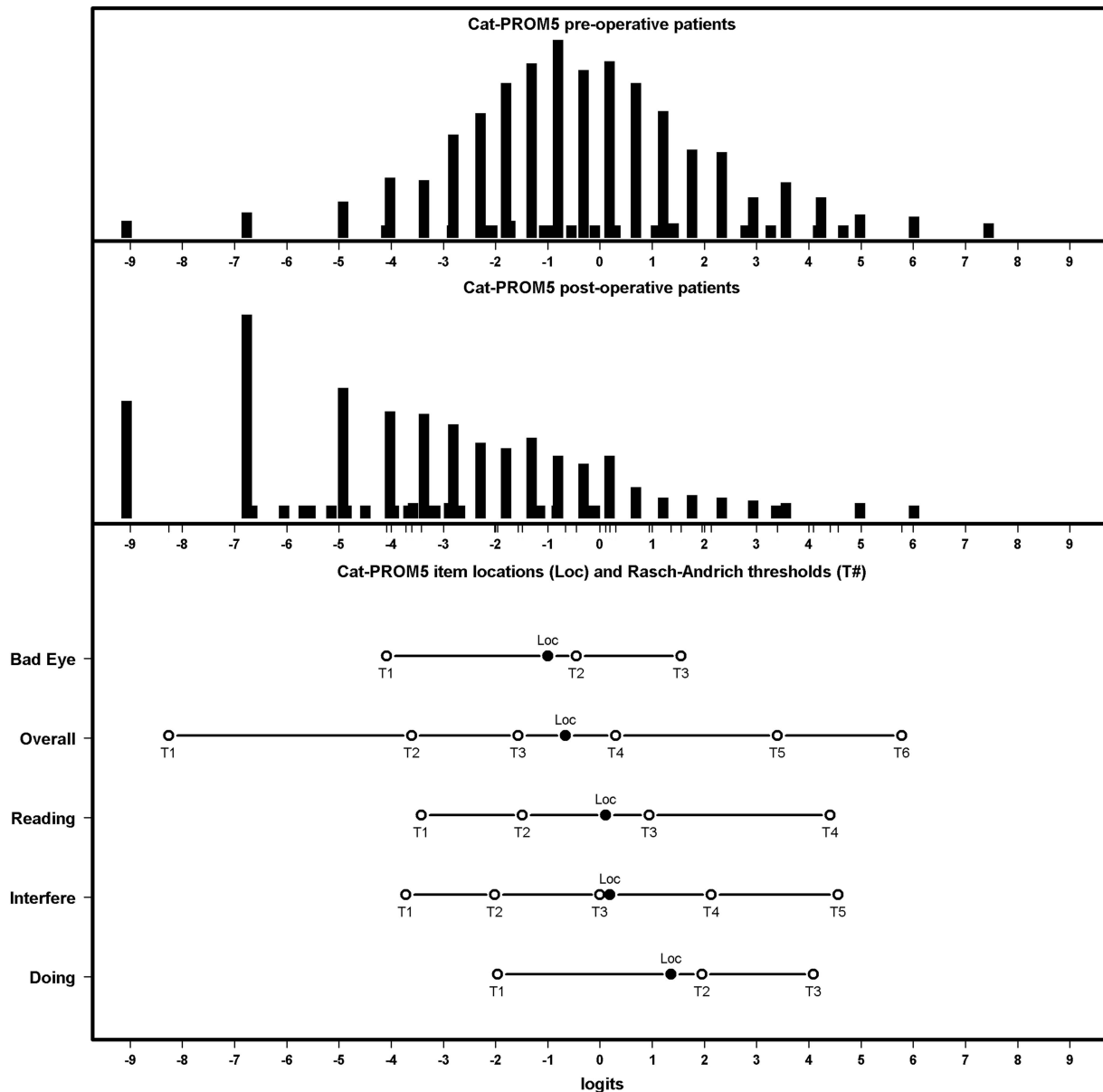


Figure 2 Cat-PROM5 Person-Item map for all cycles showing respondent distributions for pre-operative (upper panel), post-operative (middle panel) completions, and the Item Locations (Loc) and Thresholds (probability crossover points between adjacent categories, lower panel) on the same Logit scale. In total, 1266 questionnaire completions were available. Pre- and post-operative means -0.41 and -3.61 respectively.

(DIF = 0.52) each signifying a relative over-statement of visual difficulty in the presence of cataract and/or an under-statement following surgery. The third shift relating to the pre- *vs* post-operative split for the 'Doing' item just crossed into the 'significant' range. This went in the opposite direction (DIF = -0.65) implying an under-statement of the impact of visual difficulty on activities pre-operatively, which would be consistent with adaptation. Rasch Model indices for the combined data are in Table 3, all being satisfactory and confirming a well-functioning unidimensional Cat-PROM5 scale. Pearson correlation coefficients between Cat-PROM5 person measures and pre-operative LogMAR visual acuities were all highly statistically significant ($P < 0.001$) and weakly correlated: better eye 0.21; worse eye 0.19; both eyes averaged 0.24; surgery eye 0.21; fellow eye 0.14. Pearson correlation between Cat-PROM5 and Catquest-9SF²² person measures was $R = 0.85$ ($P < 0.001$; $N = 1,189$ completions).

The pre- and post-operative Cat-PROM5 means were -0.41 and -3.61 respectively with a difference of -3.20 logits and standardized effect size (Cohen's d) of -1.45 SD and -1.09 SD by the 2 methods, confirming that Cat-PROM5 is highly responsive to surgical intervention. Those pre-operative patients who had cataract affecting both eyes had a mean of +0.01 logits indicating good targeting for bilateral cataract. Small or greater, medium or greater and large or very large (0.2 SD = 0.44, 0.5 SD = 1.10, 0.8 SD = 1.76 logits) self-reported Cat-PROM5 improvements in visual difficulty were reported by 83, 72 and 68% of respondents respectively. Provided all 5 questions have been responded to, raw scores from Cat-PROM5 completions may be converted to logits using the online look-up table in Supplementary Table S3.

Discussion

A rigorous development approach to Cat-PROM5 based on Rasch and factor analysis parameters obtained from typical UK patients aged 50 years and older undergoing cataract surgery in 4 centres in England has resulted in a questionnaire with a final set of five items with robust psychometric performance. The questions are broad which allows patients to map the issues of most relevance to them to these questions, avoiding the problem of highly specific questions with limited applicability for some individuals. The set of five questions vary in presentation format, respondents thus need to consider each question individually which guards against running through the questions checking the same level for each without adequate thought to the items individually. The questions have been thoroughly piloted, have high face validity as presented, display good individual

performance indices, and the contribution of each item to the scale is highly satisfactory.

This study recruited typical patients undergoing cataract surgery who were free of other visually significant comorbidities, the intention being to avoid possible confusion of responses relating to non-cataract visual difficulties. As reported elsewhere,¹⁵ a qualitative exercise was undertaken separately with patients with both cataract and non-cataract causes of vision loss. This did not reveal serious issues with use of the questionnaire in the presence of comorbidities. Subsequent to completion of Cat-PROM5 development the questionnaire has been used in a separate group of 974 cataract patients which include the 'usual' spectrum of comorbidities. The performance of the questionnaire is similar in this group with a mean preoperative score of -0.32 logits and small or more, medium or more and large or very large improvements reported by 80%, 70%, and 62% respectively. The slightly lower proportion reporting large or very large improvements likely reflects the presence of non-cataract comorbidities.

During development, following elimination of poorly functioning, misfitting or clustering items, a unidimensional construct of visual difficulty related to cataract based on 11 items which 'move together' was established. The approach to the final item reduction used in this study included a systematic assessment of all possible alternative permutations of items following the decision to retain 2 key general questions, that is, 'VCM1_Interfere' and 'VSQ_Overall' in a final five-item set. The independent confirmatory Cycle 2 sample and the aggregated 'all cycle' analyses affirmed the psychometric performance of the final Cat-PROM5 item set. From the aggregated data, it is clear that the instrument conforms to the fundamental requirements of measurement as demonstrated by close fit with the theoretical requirements of the Rasch model (Table 3). Item invariance was satisfactory, only 3 (7.5%) fell outside of the 5% random chance limit, with DIF magnitudes borderline²³ and 2 in opposite directions, so tending to cancel each other out. Correlations with visual acuity were weak confirming that Cat-PROM5 measures a latent trait, which goes beyond traditional visual acuity. Correlation with the Catquest-9SF self-report instrument however was strong ($R = 0.85$), a direct comparison between the two instruments is published separately.¹⁵ Test-re-test repeatability was excellent (ICC = 0.89) with high responsiveness to surgical intervention for cataract and a standardised effect size, Cohen's d , of -1.45 SD (baseline SD method).

Cat-PROM5 (online S4) is offered as a well performing self-report instrument suitable for use in high volume surgical services for age related cataracts. The 'look-up table' provided in Supplementary Table S3 will allow

users to calibrate responses for their own patients and convert raw score totals from the five questions into a single measure of visual difficulty in units of logits. A fixed scoring system allows direct comparisons within and between countries though may not fully translate to other cultures and languages where a Rasch based re-calibration exercise may be required.

In conclusion, the approach used to develop Cat-PROM5 has delivered a psychometrically robust, validated, well targeted and highly responsive five-item questionnaire which can be considered as an appropriate and fit-for-purpose tool of sufficient brevity for realistic implementation in high-volume cataract surgical services.

Summary

What was known before

- Patient Reported Outcome Measures (PROMs) have been proposed for use in patients with cataract to quantify pre-operative visual difficulty and its relief from surgery.
- Despite its well-known limitations however, Visual Acuity remains the current standard for these assessments.

What this study adds

- Cat-PROM5 is a short 5-item Rasch validated scale demonstrating high-quality psychometric performance and responsiveness.
- With only five items it would be feasible to implement Cat-PROM5 into high-volume cataract surgical services as a patient centred outcome.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

JMS—overall responsibility for the study as chief investigator, conception and design of the study, obtaining funding for the study, writing and approving the manuscript; MTG—statistical analyses, writing the manuscript; NAF—overseeing study as local principal investigator, interpretation of data, reviewing and approving the manuscript; -RJ—overseeing study as local principal investigator, interpretation of the data, reviewing and approving the draft manuscript; CL—overseeing study as local principal investigator, approving the manuscript; LE—managing the study, overseeing acquisition of the data, reviewing and approving the manuscript; AL—managing the study, overseeing acquisition of the data, collection of the data, reviewing and approving the manuscript; JD—design of the study, obtaining funding for the study, reviewing and approving the manuscript.

Acknowledgements

This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Reference Number RP-PG-0611-20013). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. JLD was supported by the NIHR Collaboration for Leadership in Applied Health Research and Care (CLAHRC) West at University Hospitals Bristol NHS Foundation Trust, and is an NIHR Senior Investigator. This work was supported by National Institute for Health Research (NIHR) grant number RP-PG-0611-20013.

References

- 1 Black N, Browne J, van der Meulen J, Jamieson L, Copley L, Lewsey J. Is there overutilisation of cataract surgery in England? *Br J Ophthalmol* 2009; **93**(1): 13–17.
- 2 Rao GN, Khanna R, Payal A. The global burden of cataract. *Curr Opin Ophthalmol* 2011; **22**(1): 4–9.
- 3 Lundstrom M, Pesudovs K. Questionnaires for measuring cataract surgery outcomes. *J Cataract Refract Surg* 2011; **37**(5): 945–959.
- 4 McAlinden C, Gothwal VK, Khadka J, Wright TA, Lamoureux EL, Pesudovs K. A head-to-head comparison of 16 cataract surgery outcome questionnaires. *Ophthalmology* 2011; **118**(12): 2374–2381.
- 5 Day AC, Wormald R, Coronini-Cronberg S, Smith R. Royal College of Ophthalmologists Cataract Surgery Commissioning Guidance Development G. The Royal College of Ophthalmologists' Cataract Surgery Commissioning Guidance: executive summary. *Eye (Lond)* 2016; **30**(3): 498–502.
- 6 Cataracts in adults: management. NICE guideline [NG77]. Appendix L: Research recommendations. Available at <https://www.nice.org.uk/guidance/ng77/evidence/appendix-l-research-recommendations-pdf-167615924439>, 2017.
- 7 Gothwal VK, Wright TA, Lamoureux EL, Lundstrom M, Pesudovs K. Catquest questionnaire: re-validation in an Australian cataract population. *Clin Exp Ophthalmol* 2009; **37**(8): 785–794.
- 8 Pesudovs K, Gothwal VK, Wright T, Lamoureux EL. Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg* 2010; **36**(5): 718–732.
- 9 Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Cataract Symptom Scale: clarifying measurement. *Br J Ophthalmol* 2009; **93**(12): 1652–1656.
- 10 Wright BD, Bell SR. Item banks: what, why, how. *J Educ Meas* 1984; **21**: 331–345.
- 11 Hahn EA, Cella D, Bode RK, Gershon R, Lai JS. Item banks and their potential applications to health status assessment in diverse populations. *Med Care* 2006; **44**(11 Suppl 3): S189–S197.
- 12 Donovan JL, Brookes ST, Laidlaw DA, Hopper CD, Sparrow JM, Peters TJ. The development and validation of a questionnaire to assess visual symptoms/dysfunction and impact on quality of life in cataract patients: the Visual

- Symptoms and Quality of life (VSQ) Questionnaire. *Ophthalmic Epidemiol* 2003; **10**(1): 49–65.
- 13 Laidlaw DA, Harrad RA, Hopper CD, Whitaker A, Donovan JL, Brookes ST *et al*. Randomised trial of effectiveness of second eye cataract surgery. *Lancet* 1998; **352**(9132): 925–929.
 - 14 Frost NA, Sparrow JM, Durant JS, Donovan JL, Peters TJ, Brookes ST. Development of a questionnaire for measurement of vision-related quality of life. *Ophthalmic Epidemiol* 1998; **5**(4): 185–210.
 - 15 Sparrow JM, Grzeda MT, Frost NA, Johnston RL, Liu CSC, Edwards L *et al*. Cataract Surgery Patient Reported Outcome Measures: A head-to-head comparison of the psychometric performance and patient acceptability of the Cat-PROM5 and Catquest 9SF self-report questionnaires. *Eye* 2018; epub ahead of print 26 January 2018; doi:10.1038/eye.2017.297.
 - 16 Andrich D. Rasch Models for Measurement: SAGE Publications Inc.: Thousand Oaks, CA, USA, 1988.
 - 17 Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med* 2003; **35**(3): 105–115.
 - 18 Solari A, Grzeda M, Giordano A, Mattarozzi K, D'Alessandro R, Simone A *et al*. Use of Rasch analysis to refine a patient-reported questionnaire on satisfaction with communication of the multiple sclerosis diagnosis. *Mult Scler* 2014; **20**(9): 1224–1233.
 - 19 Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989; **70**(12): 857–860.
 - 20 Wright BD, Masters GN. *Rating Scale Analysis*. Chicago: Mesa Press, 1982.
 - 21 Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. *Mahawah*. New Jersey: Lawrence Erlbaum Associates, 1988.
 - 22 Lundstrom M, Pesudovs K. Catquest-9SF patient outcomes questionnaire: nine-item short-form Rasch-scaled revision of the Catquest questionnaire. *J Cataract Refract Surg* 2009; **35**(3): 504–513.
 - 23 Wilson M. Constructing measures. *Mahawah*. New Jersey; London: Lawrence Erlbaum Associates, 2005.
 - 24 Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans* 1994; **8**: 370.
 - 25 Wright BD. Local dependency, correlations and principal components. *Rasch Meas Trans* 1996; **1996**(10): 509–511.
 - 26 Kim J-O, Mueller CW. Factor Analysis. *Statistical Methods and Practical Issues*. Newbury Park, London, New Dehli: Sage Publications, 1978.
 - 27 Steiger JH, Lind J. *Statistically-based tests for the number of common factors*. Conference Paper. Iowa City, 1980.
 - 28 Hu L-T, Bentler P. Evaluating Model Fit. In: Hoyle RH ed. *Structural Equation Modeling: Concepts, issues, and applications*. Thousand Oaks, California: Sage Publications, 1995; 76–99.
 - 29 Yu C-Y. *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes*. Doctoral Dissertation, University of California, 2002.
 - 30 Brown TA. *Confirmatory Factor Analysis*. New York, London: The Guilford Press, 2006.
 - 31 Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990; **107**: 238–246.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2018

Supplementary Information accompanies this paper on Eye website (<http://www.nature.com/eye>)